CAPRi: Cloud-based Analytic Framework for Precipitation Research

Presenter: John Beck

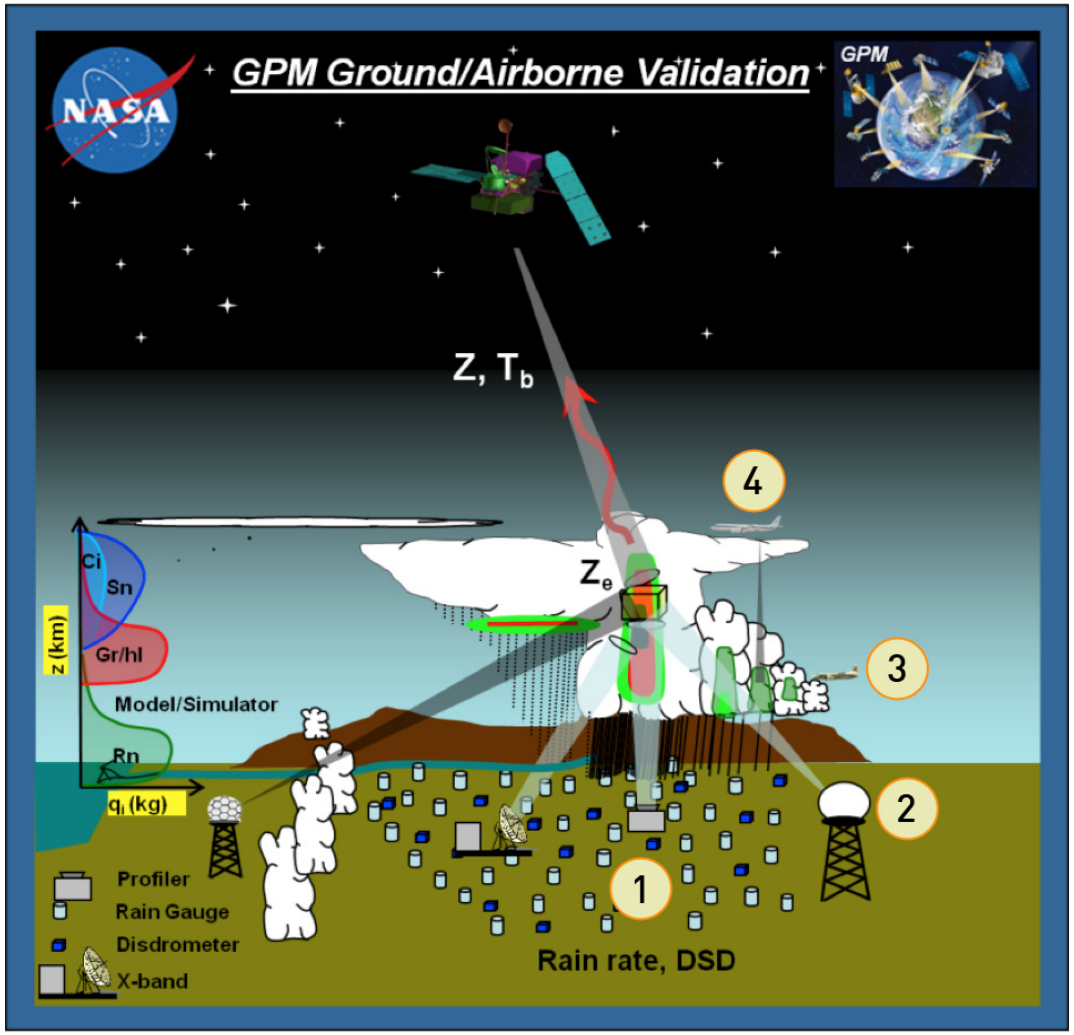
PI: John Beck

Team Members: Todd Berendes, Patrick Gatlin, Anita LeRoy, Geoffrey Stano

Program: AIST-18

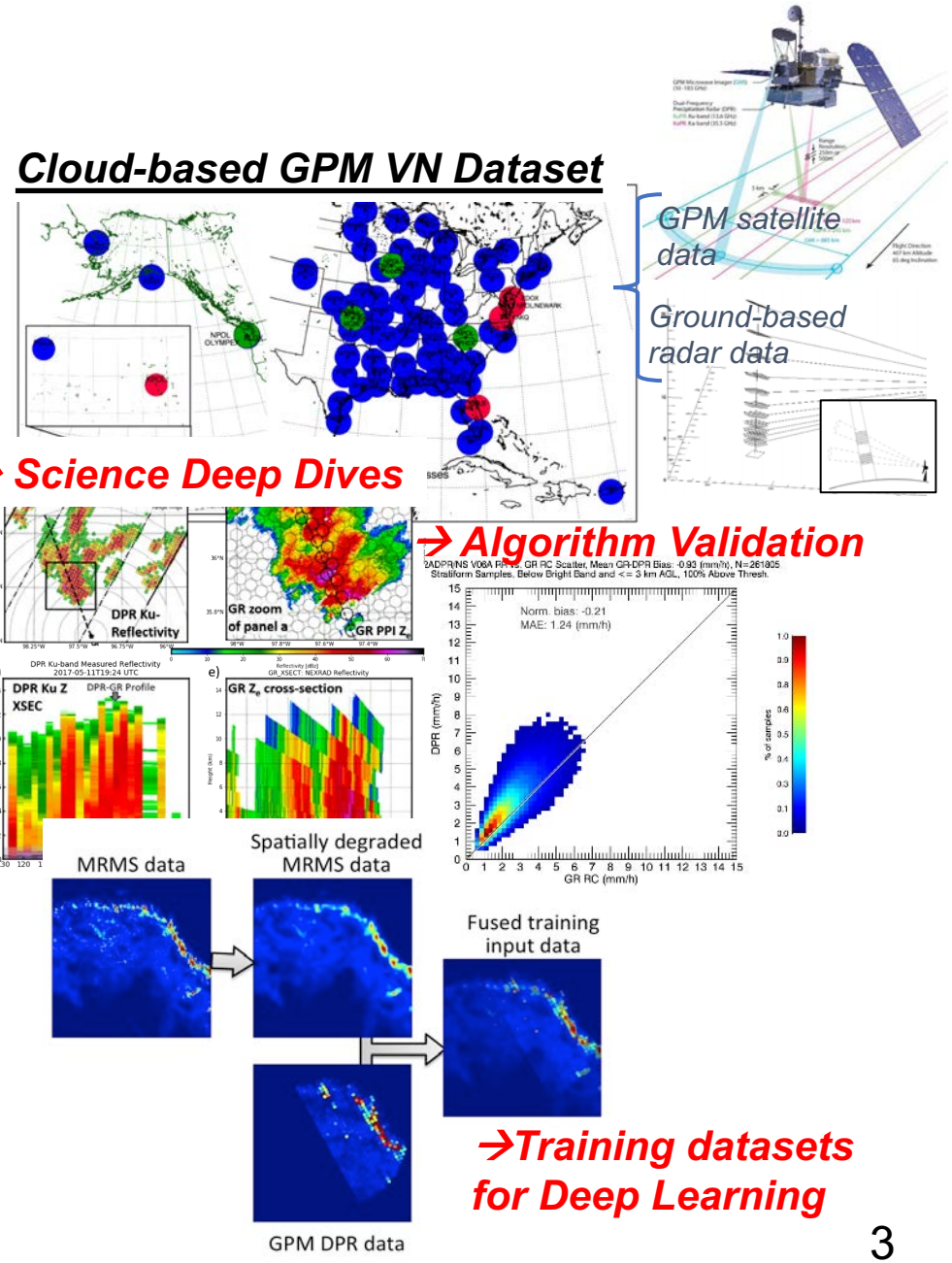
Problem to Solve

- Assembling and preparing large Earth science datasets often requires a great deal of time and effort, especially those requiring analysis of field campaigns.
- Many scientific data sets and analysis tools are now located on cloud-based platforms that can have steep learning curves.
- The complexity of such datasets limits their use in research and applications, especially advanced data analytics and data fusion techniques such as deep learning.



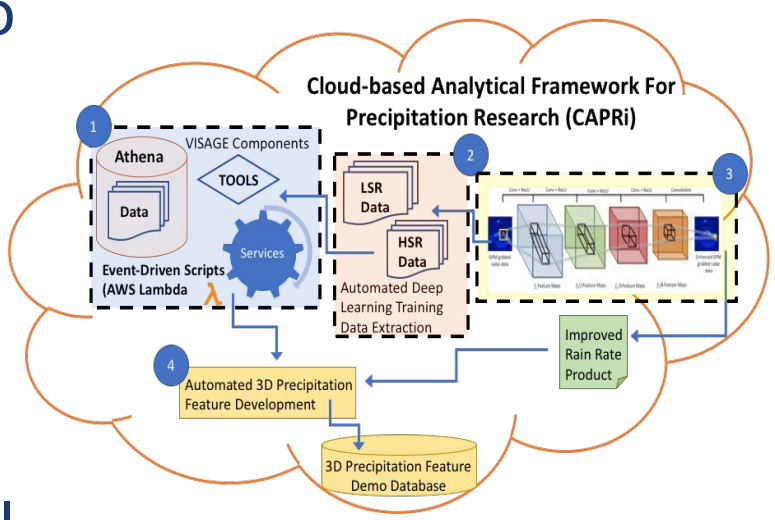
Solution

- Develop a **C**loud-based **A**nalytical **F**ramework for **P**recipitation **R**esearch (CAPRi) providing users with tools for on-demand data querying, fusion, sub-setting, extraction, visualization, and analysis of precipitation data.
- Integrate Deep Learning models into CAPRi for super-resolution of Global Precipitation Measurement (GPM) Dual Precipitation Radar (DPR) data.
- Use CAPRi services and the results from the super-resolution Deep Learning exercise to identify 3D precipitation features and create a demonstration database of the results.



Technical Details (Analytical Framework)

- CAPRi will leverage cloud-native technologies to develop an architecture hosting NASA's space-based Global Precipitation Measurement Validation Network (GPM VN) data integrated with Deep Learning models to provide an analysis-optimized cloud data store with access via on-demand cloud-based serverless tools.
- This approach provides flexible scaling to support on-demand data discovery, rendering, and analytics at minimal cost. One challenge will be the shear volume and file size of the data for serverless technologies to handle.
- Serverless computing tools from Amazon Web Services (AWS), including step and lambda functions, S3 object storage, and the Athena stateless query service will be used to extract, prepare, and provide GPM data for Deep Learning analysis.



Technical Details (Deep Learning)

- Deep Learning is a very active field and many cloud-based tools exist. However, these tools do not provide support for custom querying, pre-processing and organization of large volume training and/or test data.
- Since, most of the mainstream Deep Learning tools are designed to work with images or structured data formats, they require data to be reformatted into specific formats. As is often the case for many scientific data sets, precipitation data is not in a Deep Learning ready data format, requiring pre-processing and reformatting.
- CAPRi seeks to ease the burden of data conversion and formatting by providing a customizable framework that produces analysis-ready data and provides a cloud-based interface for data access, analysis, and creation of training and testing features in a supported format.

Technical Details (Precipitation Features)

- To demonstrate this technology, we propose to use CAPRi services for identifying convective scale 3D precipitation features.
- Convective scale allows us to better separate out trends in how precipitation is being delivered. The 3D component will allow us to better investigate the drivers, link to lightning, and support future modeling efforts.
- We will integrate a Deep Learning architecture into CAPRi to conduct super-resolution on GPM precipitation data using Multiple Radar Multiple Sensor (MRMS) data as a high-resolution ground truth.
- Results will be used within an advanced Deep Learning Model for identifying 3D precipitation features.